AI Simplified # Discriminative AI-Chapter 3

#Artificial Narrow intelligence

RECAP :

- Scalar, Vector , and matrix
- Types of Probability

ML (Machine Learning)

- Liner regression

Data :

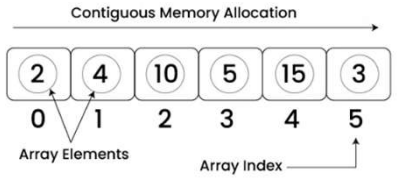
- Structure Data
- Semi structure Data
- Text/image/Video/Audio
- Data format
- EDA

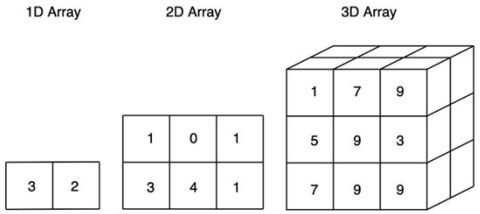
Subset of AI is every where (in a way or other)

- Spam email filter
- Price prediction
- Sentimental analysis
- Autonomous vehicle
- Speech recognition
- Spell check and auto correct
- Language translation
- Word auto suggestion
- Transaction verification
- Traffic management
- Hand written digit recognition
- Interactive chat agent
- Object identification
- Music generation
- Recommendation systems
- Netflix/Prime movie suggestion
- Earthquake prediction
- Climate change prediction
- Disaster recovery management
- Insurance verification and approval
- Cancer prediction

.....and many more

Recap - Mathematics (liner algebra)





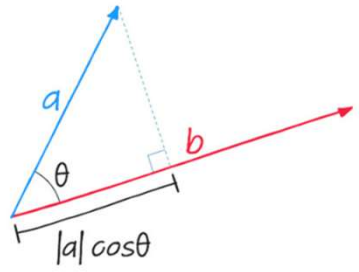

```
# 1D array
arr = [1, 2, 3]

# 2D array
arr_2d = [[1, 2, 3], [4, 5, 6]]

# 3D array
arr_3d = [[[1, 2], [3, 4]], [[5, 6], [7, 8]]]
```

Array

- N dimensions
- Ordered collections of elements



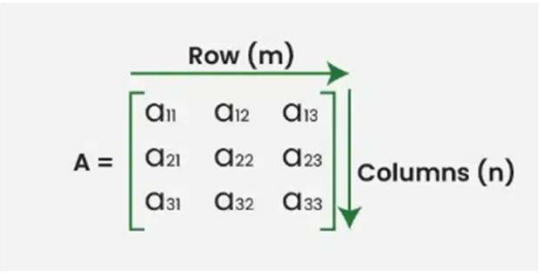
$$a \cdot b = |a|\,|b| \cos\theta$$

$$\vec{A} = 5\hat{\imath} + 8\hat{\jmath}$$

$$\vec{B} = 1\hat{\imath} - 3\hat{\jmath}$$

Vector

- One Dimensional Array
- Has both magnitude and direction



$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \quad B = \begin{bmatrix} 4 & 5 \\ 6 & 7 \end{bmatrix}$$

Matrix

- Two Dimensional in rows and column

Recap-Probability "The world of uncertainty "

- Inferential statistics : using smaller sample to infer information for larger population
- Probability distribution : Distribution of probability for each outcome of an experiment
- Probability types
  - Discrete Probability : Random events that have a finite or count of possible outcomes
    - Uniform distribution : All outcome has **equal probability**
      Ex : Rolling a dice

    - Binominal distribution : No of success in fixed number of independent trials , each with same P(X)
      Ex: Toss a coin 100 times and to note how many times tail comes

    - Cumulative distribution: P(event 1 OR event 2) = P(event 1) + P (event 2)
      Ex: Out of 100 random people how many are more than 21 years.

    - P(event 1 AND event 2) = P(event 1) X P (event 2)

$$\text{Probability of an event happening} = \frac{\textit{Number of favorable outcome}}{\textbf{Total number of outcomes}}$$

About ME

Trying to make this world a better place to live

AMBASSADOR

Gartner
Peer Community

COSCUP
開 源 人 年 會

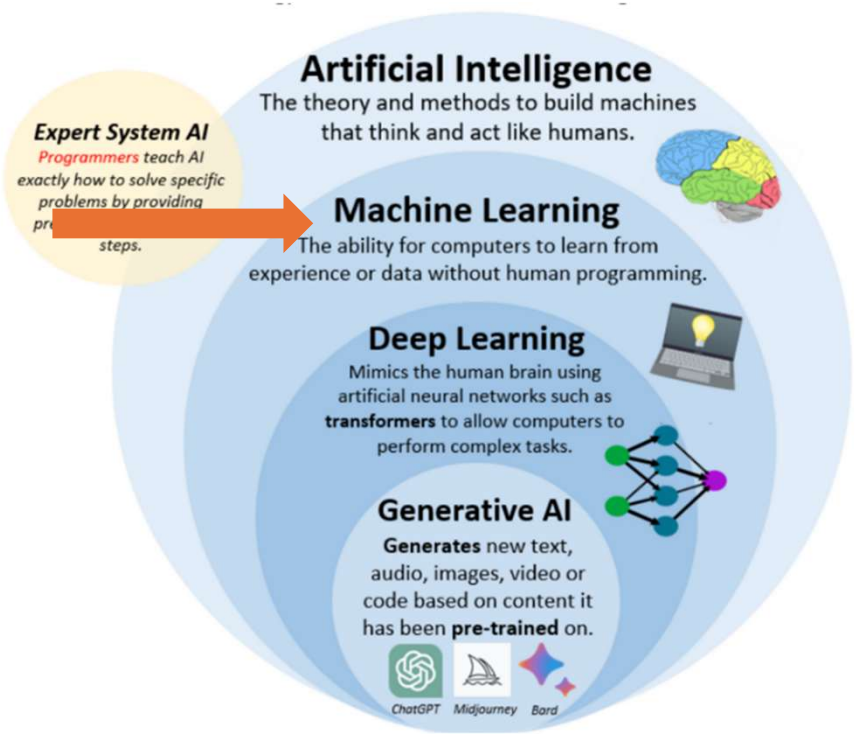
AI Simplified

AWSSimplified

Preet Sharma

Problem Solver, Solutions Architect | Helping Organizations with Value-Innovation Solutions | C...

ML



Machine Learning (ML)

ML is an approach to learn complex pattern from existing data and use this pattern to make predict on unseen data.

- Learn
- Pattern
- Data
- Predict
- Unseen

Machine Learning

Types of ML algorithms :

- Regression (Supervised learning)
- Classification (Supervised learning)
- Clustering (Unsupervised learning)

Q1: You have the past data of two cricket teams on the performance of the teams based on different parameters and the match results. You have to predict which team will win. Is it supervised or unsupervised ?

Q2: You feed a large collection of spam emails to the learning model to identify the different sub-groups of these spam mails. No labels are presents in the data set. Is it supervised or unsupervised ?

Q3: Consider a large data set of the medical profiles of cancer patients. This data contains no labels for the medical profiles of the cancer patients. The model has to learn whether there might be different groups of such patients for which separate treatments might be tailored. Which ML problem is this ?

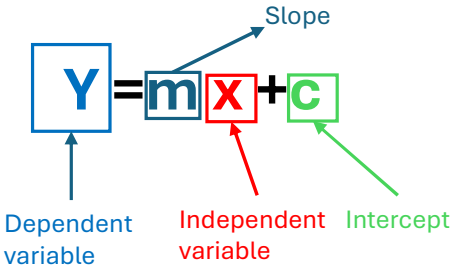
ML Use cases (Specific to Phone only)

- You open your phone using finger print or face authentication. Its ML
- You type something , you see auto suggestion and auto correction . Its ML
- You click photo and enhance it . Its ML
- You use your voice for siri or google assistant. Its ML
- You translate one language to another. Its ML
- Phone Camera detect face and do auto focus. Its ML
- You see movie recommendation on Netflix or Prime video. Its ML
- Phone take speech as input and generate text. Its ML

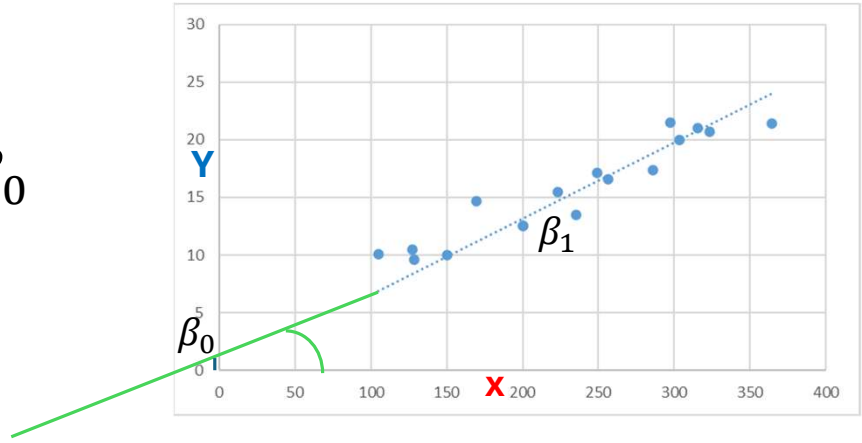
Machine Learning (Simple liner regression)

Sample Application of Liner regression

- Predicting cab ridership (demand) based on time and week of the day, weather conditions, holiday or not, traffic conditions, etc.
- Predicting fares based on route, time of day and week, ride type, number of available drivers, tolls, etc.
- Predicting ride duration.
- Predicting surge prices based on demand and offering discounted prices during off-peak times, etc.
- Predicting dynamic flight fares based on destinations, time left to departure, weather conditions, etc.
- Predicting bank performance indicators like return on asset (ROA), Net Interest Margin (NIM), etc., based on bank specific variables like asset quality, capital adequacy, liquidity management (measured using certain ratios) and macroeconomic variables like GDP growth rate, inflation rate, etc.
- Predicting length of stay (LOS) in the hospital based on age, disease type, severity, etc.
- Predicting sales in an eCommerce company based on the social media ad spends, SEO spends, etc.
- Predicting profits in a retail chain based on the size of the city, demographics in the locality, etc.
- Predicting wafer quality in semiconductor manufacturing based on temperature, pressure, gas flow, etc.



$$\hat{y} = \beta_1 x + \beta_0$$



Machine Learning (Simple liner regression)

Understand relationship between attributes

- As per IATA (in 2012), about 33% of airlines operating costs are due to fuel consumption.* Fuel burn is dependent on landing weight.** IATA defines cost of weight (COW) as: 'extra fuel burn, as a function of additional weight'.

- IATA calculates the COW factor by looking at the relationship between landing weights and the corresponding fuel burns per hour. Ref: Guidance Material and Best Practices for Fuel and Environmental Management, 5th edition, October 2011, p34

https://www.iata.org/pressroom/pr/Pages/2013-07-16-01.aspx
https://www.aircraftit.com/Operations/eJournals/eJournal/Spring-2012/Reviews/A-New-Approach-to-Cost-of-Weight-COW.aspx

Landing Weight (kg)	Fuel Flow (kg/h)
146167	4700
148500	4800
150333	4700
159000	5100
159333	5325
159667	5025
160167	5300
160333	5050
160583	5250
160833	5225
161167	5075
161667	5237.5
162583	5087.5
162833	5125
164167	5150
164500	5400
165000	5375
165000	5150
165333	5375
165333	5075
166000	5350
166583	5087.5
166750	5375
167167	5500
168917	5575
169250	5350
169500	5650
169667	5675
170000	5562.5
170167	5662.5
170667	5550
170750	5655
170833	5375
170833	5652.5
170917	5375
173167	5662.5
175000	5550

Lets Make Liner regression model : ⟶

Machine Learning (Simple liner regression)

CRISP –DM Framework (Cross industry standard process for Data mining)

1. Business understanding
2. Data understanding
3. Data Preparation
4. Data Modelling
5. Model Evaluation
6. Model Deployment

EDA (Exploratory Data Analysis)

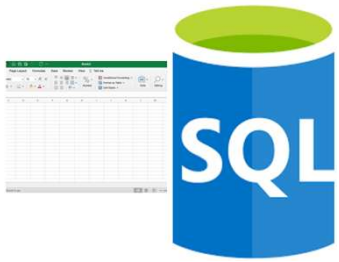
- Central tendency of data
- Fix rows and column
- Missing values
- Standardize numbers
- Standardize text
- Fix invalid values
- Filter data

DATA

Thing that gives direction to world.

Data Types

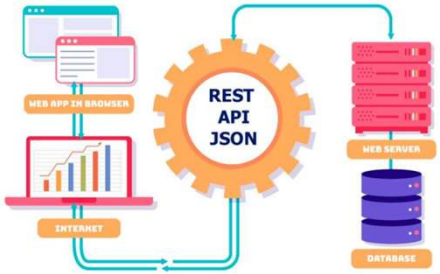
Structure DATA



- Information organized in a predefined, consistent format, making it easy to store, retrieve, and analyze

01 | Spreadsheets, CSV
02 | Enterprise Data Warehouses
03 | ERP e.g. SAP, Oracle, Zoho, Tally
04 | CRM e.g. Salesforce, Zendesk, Adobe
05 | HR mgmt. eg. Workday, Ceridian, Oracle

Semi Structure DATA



- Type of data that falls between structured and unstructured data.

01 | Data exchange between server and client
02 | Configuration files in software applications
03 | Application log files info about events, errors
04 | Used for encoding geographical data e.g. GeoJSON
05 | IoT devices exchange data between sensors, devices, and servers.

Unstructure DATA

- No Predefined data model or schema

01 | Audio , Video and any other media files
02 | Emails
03 | IoT sensor log data
04 | social media data
05 | image , satellite data

Data Types

Unstructured Data

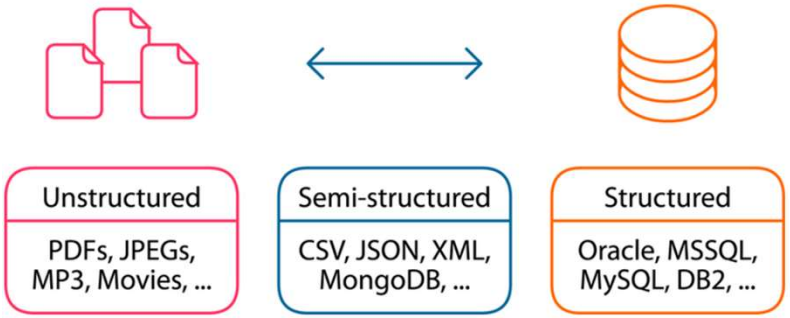
The university has 5600 students. Shaun (ID Number: 160801), 18 years old Communication study. Linh with ID number 160802, majoring in Accounting and is 20 years old. Ahmed from Psychology study program, 19 years old, ID number 160803.

Semi-Structured Data

```
<University>
 <ID Number="160801">
  <Name="Shaun">
  <Age="18">
  <Program="Communication">
 <ID Number="160802">
  <Name="Linh">
  <Age="20">
  <Program="Accounting">
..... </University>
```

Structured Data

ID	Name	Age	Program
160801	Shaun	18	Communication
160802	Linh	20	Accounting
160803	Ahmed	19	Psychology



- In Data Science, ML algorithms are typically applied to structured data.
- In AI, ML algorithms are typically applied to text, video, image, audio
- Real world deployments often rely on a pipeline of data pre-processing, data transformations, post-processing

What will be Helpful ?